Use of Administrative Data in Czech Business Statistics
(Seminar Using Administrative Data in the Production of Business Statistics
– Member States experiences)

1.   Reasons for use of administrative data

The requirements on range of statistical information, as well as their quality, are increasing in several last years. On the other hand respondents are not willing to provide necessary data for statistical surveys. The State administration imposes pressure to decrease administrative burden of respondents. Moreover financial resources of the Czech Statistical Office are limited. Number of employees decreased in 2008.
One of possible solutions of the situation is to increase use of administrative data.

2.   Main administrative data sources used in Business Statistics

2.1   The Commerce Register
The owner of the source is the Ministry of Justice.
Contents of data source:
- The legal persons have an obligation to register at the Commerce Courts (7 places) according to the Commercial Code.
-    The file with changes in the register is provided to the Czech Statistical Office in daily periodicity.
-    We get information about birth and death of enterprises, their activity, address etc.

2.2   The Trade Register
The owner of the source is the Ministry of Industry and Trade.
Contents of data source:
- The natural persons have an obligation to register at the Trade Offices (500 places).
-    The file with changes in the register is provided to the Czech Statistical Office in monthly periodicity.
-    We get information about birth and death of enterprises, their activity, address etc.

2.3   Czech Social Security Administration (CSSA)
The owner of the source is the Ministry of Labour and Social Affairs.
Contents of data source:
-    Enterprises with employees have an obligation to register for the payment of social security. That is why database comprise almost all enterprises with employees. It is now approximately 265 000 enterprises.
-    The register contains number of insured people for each company (employee with full-time, part-time contract)
-    The source comprises information about division units into regions, too.
-    We get files with data monthly, with 2 month delay after reference period
-    We don't obtain updates or corrections, just delayed records every months
-    Comparable data are available from 2002 to 2008 and than from 1/2009 to 12/2009
Key variables:
- number of social security policyholders (aggregates by enterprises)
- assessment of social security insurance (proxy variable for wages and salaries)

Problems with the source:
- We get data in 2 files. The first of them contains number of social security policyholders and the second assessment of social security insurance. The department, which occupies with administrative data, joins data in these two files. Unfortunately not all records match.
- The Czech Statistical Office uses in the Business Register as well as in all surveys the unique identification number. CSSA uses a different identification within their sources. The unique identification number is mostly included in the file with number of social security policyholders, but sometimes it is missing. That is why we aren't able to use all obtained records.
- The enterprise can pay social security insurance for more enterprises, with which is connected. This problem is solved by special database containing a list of connected companies.
- Due to changes in the legislation and extension of CSSA duties delivery of data was delayed at the beginning of 2009. The quality was insufficient.

Checks and corrections of administrative data:
- Before we use data in further process, we do some checks and corrections. Administrative data are compared with data sent in statistical questionnaires each month.
- New administrative data of an enterprise are compared with three months average counted from previous administrative data. This is done for both variables.
- If new value of one of variables, the number of social security policyholders or the assessment of social security insurance, is much different from previous values in time series, we review whether the value of the another variable changed in correspondent way.
- In case we find substantial differences and we consider administrative data unreliable, we do manual corrections.

2.4   Income tax returns
The owner of the source is the Ministry of Finance.
Contents of data source
- All enterprises have an obligation to pay Income tax. That is why the source should contents almost all active enterprises (both natural and legal persons). The source covers as well small enterprises.

- The data are annual. We obtain preliminary data with 9 months delay after reference period it means, in September and final data with 12 months delay. Updates of preliminary data are sent with final data.
- Complete data including balance and income statement are available from 2006, for 2005 we get reduced data set.

Key variables:
- sales revenues, sales of sold goods, margin, value added
- wages, number of employees
- structure of assets, tangibles, intangibles, liquid assets, stocks
- structure of liabilities and equity
- financial costs and revenues etc.

Problems with the source:
- The Ministry of Finance uses a different identification, no the unique identification number. That is why we have to transfer the Ministry of Finance's identification to the unique identification number.

- Data are available with long delay (9 months)

Checks and corrections:
- We try to identify whether there isn't order of magnitude error in data. This error would mean an enterprise handed in the data either in thousands of Czech crowns or halers, while it should hand in data in Czech crowns.
- We compare the current data on Income Tax Return of an enterprise with last year's data. Ratio is counted for each variable. If ratio indicates order of magnitude error, the value is divided by ratio.
- VAT data and data from survey are used for similar checks.

2.5   VAT (value added tax) Tax Returns

The owner of the source is the Ministry of Finance.

Contents of data source
- The enterprise has an obligation to pay value added tax in case that its turnover in last 12 months exceeds 40 000 Euro (1 Mil. CZK). Whether enterprise has to pay VAT quarterly or monthly depends on the size of turnover. If the turnover is between 40 000 and 400 000 Euro the enterprise pays quarterly. If the turnover in last 12 months exceeds 400 000 Euro (10 Mil. CZK) the enterprise has to pay VAT monthly.
- At present 100 000 enterprises pays VAT monthly and 300 000 enterprises quarterly.
- We obtain data with 2 months delay.
- Updates and corrections of data are sent every month.
- Data are available from January 2005.
- It contains only part of enterprises because a lot of companies don't have to pay VAT tax.

Key variables:
- domestic turnover
- export/import of goods EU countries/outside EU
- imports of services from EU/outside EU

Problems with the source:
- Due to reorganization of information system of the Ministry of Finance, delivery of data was 5 months delayed at the beginning of 2009.

Checks, corrections and processing:
- When data for new period are received, two indicators (turnovers) are calculated from variables on VAT Returns for every enterprise. Each of turnovers is calculated from different variables. Both turnovers are compared in time with twelve previous values for every enterprise. Program calculates standard deviations and marks data, which are outside confidence interval. The result indicator – VAT turnover, which is used in the following statistical processes, is set as calculated turnover 1. If turnover 1 is marked (outside confidence interval), turnover 2 is used. If both turnovers are marked, we consider them as correct and turnover 1 is used.

3. How we use administrative data

Until 2008 the administrative data were used mainly for the Business Register updating. The number of social security policyholders and the turnover from VAT were used for set of a calibration tables.

From 2009 statistical surveys within the Business Statistics were reformed. We still use statistical surveys as a main source of information. The Monthly survey serves for providing of the requirements of the STS regulation. The Quarterly survey is a main source for quarterly GDP estimation. The Annual survey helps us to ensure the data according to the SBS regulation.

In order to decrease an administrative burden of small enterprises, we do so called a negative coordination. It means if an enterprise is drawn for a monthly survey, it cannot be drawn for any of quarterly nor for any of annual surveys, unless it belongs to a stratum with 100% sample rate. There are several mutations of the questionnaires used within the quarterly survey in order to ensure enterprise report the information on one indicator only once. If an enterprise is drawn for the monthly as well as for the quarterly survey (it belongs to the stratum with 100% sample rate), the information, for example, on the number of employees reported in monthly questionnaires of the respective quarter is used within the quarterly survey as well. The administrative data are more used in the imputations in our new system. The use of administrative data goes through all process of statistical surveys.

3.1 Business Register updating

Data in the BR are updated from various sources. The basic information about birth or death of an enterprise comes from the Commerce Register or the Trade Register. Furthermore several information are updated:

Information about activity/inactivity: As far as we find some information about an enterprise in the administrative source, it means the enterprise pays the social security insurance for it's employees or it declares the Income Tax Returns or the Value Added Tax Returns, we consider it as active.

Number of social security policyholders: We update two items in the Business Register by number of social security policyholders for each enterprise in December. The first item presents annual level and is filled with maximum value from the initial value (January), the final value (September) and the average value of the number of policyholders in the relevant year (numbers for October to December are estimated). The second item presents short-term level and is filled with average value of number of policyholders for last three known months. The short-term level is updated during a year, the updating occurs only if the change of value would be considerable (difference more than 50).

Turnover, assets, financial assets, tangible property, intangible property: These items are updated from Income Tax Returns of previous year once a year, in December.

VAT Turnover: This item is updated form the Value Added Tax Returns in December. At that time we have data from January till September. The annual value is counted as sum of all 12 months, where values for October to December are estimated, frequently by application of the September data. Moreover two irregular updates of item carried out in 2009. We would like to execute updating regularly in future.

3.2 Frames and Samples

The frames and the samples of the statistical surveys within the Business Statistics are designed from the Business Register. The more recent data in the Business Register, the better samples we have. The annual level of number of social security policyholders is used for sampling of annual survey (SBS), the way of filling ensures that we don't miss important units in the sample of annual survey.

As mentioned above, some information in the Business Register are updated during a year (for example number of policyholders), but only in case that change of value is significant. It

is because frames and samples of STS are updated during a year and the goal is to register significant changes, but at the same time to ensure steadiness.

As regards annual surveys, frames and samples are drawn on the base of previous year information. When administrative data for relevant period are available, frames are updated.

3.3 Imputations

From 2009 the Czech Statistical Office introduces a new method of processing of statistical surveys, we gave up grossing up and we focus on imputations of individual enterprises in the frame. The smallest enterprises aren't imputed individually, but as a part of a group with the same characteristics.

For each indicator on questionnaire is specified auxiliary variable from administrative source, eventually several auxiliary variables. Values of auxiliary variables (for example number of social security policyholders, turnover) are known for all enterprises in frame. Frames and samples are stratified by main activity (NACE classification), size of enterprise (number of social security policyholders) and for some surveys we use other characteristics. Coefficient, which reflects relation between estimated indicator and auxiliary variable, is calculated for each strata.

Outliers are eliminated from the calculation. The calculated coefficient is applied for non-respond enterprises and enterprises, which aren't in the sample. The value of their indicator is calculated by multiplication of their auxiliary variable by coefficient. In case that in strata isn't enough respond enterprises, coefficient is calculated from wider strata.

Example: Imputation of sales of sold goods with aid of turnover VAT
Coefficient of strata is calculated from respond enterprises that aren't outliers:
Coefficient of strata = (sum of strata (weight* sales of sold goods)) / (sum of strata (weight * VAT turnover))
Sales of sold goods for every enterprise in frame (or group of enterprise) is calculated according to the formula:
Sales of sold goods of enterprise = Coefficient of strata * VAT turnover of enterprise

All variables for individual enterprise are calculated by mentioned method. There is relation between some indicators on questionnaire. When variables are calculated separately, relations could be broken. That is why we do corrections after imputations.

Example: correction of sales
Sales = Revenues from services + Sales of goods + Sales of own product
1000 = 300 + 300 + 420   Variables were imputed separately
0   = 0.05 + 0.2 + 2.0   variance of variables in strata
1000 = 299 + 295 + 406   correction
The greater is the variance, the greater is the size of adjustment of the value.

3.4 Advantages of administrative data use
- It enables to identify an economically important company (e.g. enterprise with high number of policyholders, enterprise with low number of employees, but high assets)
- Enterprises are monitored during the year. Samples of annual as well as short-term surveys could be updated. The enterprises with big difference in number of policyholders at the beginning and at the end of year aren't missing in the sample of annual survey.
- Reduction in number of surveyed indicators (especially within annual surveys).
- Reduction in number of sampled enterprises.

3.5 Drawbacks of administrative data use
- Statistics are dependent on an administrative data sources
- Legislative amendments could have an impact on availability, form and quality of the administrative data (In Czech legislative isn't confirmed an obligation to consult data requirements with the Czech Statistical Office.)
- The unique identification number, which is used in the Business Register and in statistical surveys, doesn't have to be used as identification in administrative data (If this case, the conversion must be done.)
- The indicators from statistical questionnaires aren't methodically identical with the indicators from administrative sources (e.g. the number of social security policyholders isn't identical with the number of employees. In fact number of policyholders is higher.)
- We don't obtain updates and corrections from some administrative sources (CSSA)
- The administrative data are available with quite long delay
- If we find an unreliable data, we can't contact the enterprise (respondent) with questions about correctness. We have to solve the situation through the state institution, which provides us with the administrative data.
- In some cases, a state institution hires a private company for processing of an administrative data. For this situation is typical that negotiations on correcting errors in data are complicated and lengthy.

4. Future plans in the field of administrative data

Despite the drawbacks of the administrative data use, the work with them is beneficial to the statistics and we plan to continue to use them.

As we know, by the end of June 90% of natural persons handed in the Income Tax Return that is why we try to gain access to these data with shorter delay.

We would like to explore feasibility of complete replacement of annual survey of natural persons by available administrative data

Used sources:
- Present and Future Role of Administrative Data in Structural Business Statistics - by Ondřej Vozár
- A Reform of Statistical Surveys within Business Statistics in conditions of the Czech Statistical Office – by Juraj Lojka
- Central processing of SBS - by Pavel Skotnica
- Tax Administrative Data Sources – by Radim Petr

Prepared by: Jana Gotvaldová